FOS course

Molecular Data Science

In person teaching

- Key to education but now a fragile luxury
- Test if you have symptoms – do not take risks
- Hybrid teaching: Teams
  (have buddy on Discord?)
- Be flexible & mild for yourself and each other


Experimental Biologist
Computational Biologist
NO EATING
NO DRINKING
NO BREATHING
NO OPEN TOE SHOES
NO SHORTS
NO MINI SKIRTS
ALWAYS LAB COATS ON
LIFE IS BEAUTIFUL


pipetman

Molecular Data Science

Solving a biological puzzles without touching a pipette

Data drive the modern life sciences

“Tomorrow’s discoveries will be made by today’s early-career researchers. ... We also hope to be guided by them, so that we can meet their needs when it comes to publishing their work, as research becomes more data-rich and computationally heavy.”

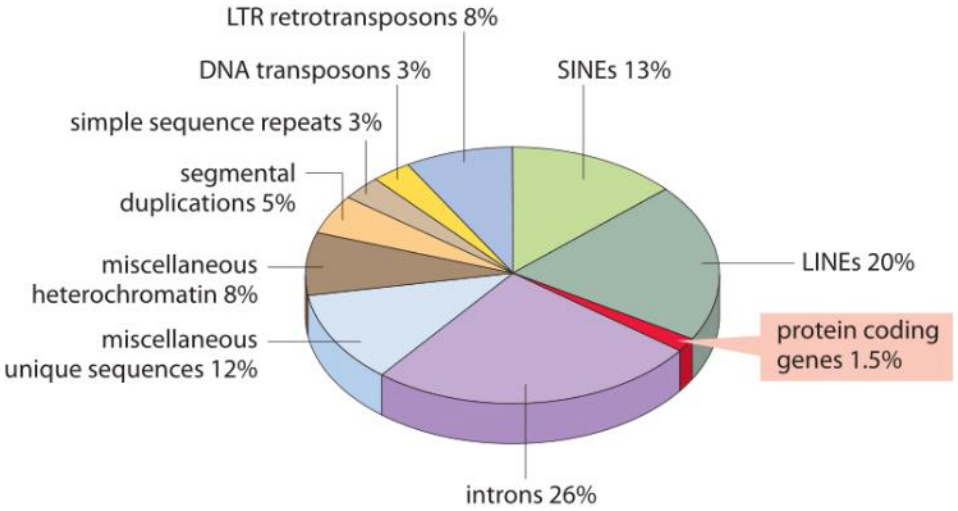
Magdalena Skipper,
Editor-in-Chief Nature

Many discoveries remain

The human genome

- ~ 3.2 billion bases (billion = miljard)
- ~ 22 thousand protein coding genes
- ~ 98% of human genome is non-coding

… and we hardly have a clue.



Many discoveries remain

Our human DNA is uncharted territory

- Dedicated paper on 5400 genes.
- 90% of papers just on 2,000 genes.
- 2015: 50% of research devoted on 3,000 genes known by 1991

Stoeger et al. PLoS Biology 2018

Features of current biology

Data intensive

- Acknowledge lack of knowledge: genome-wide

Gene

PCR

Chosen by researcher

Genome-wide

Array

All current knowledge

Whole genome

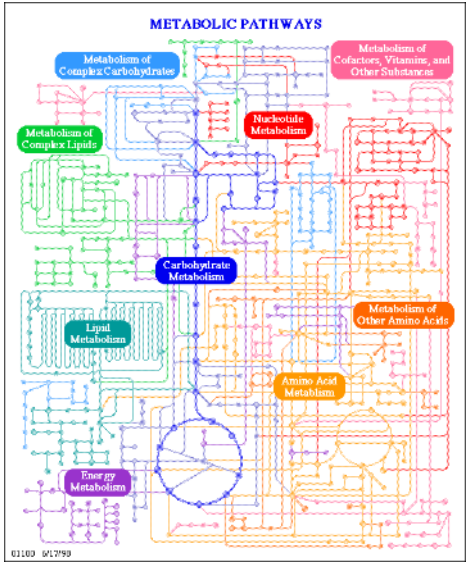
Sequencing

Everything there is

Features of current biology

Data intensive

- Acknowledge lack of knowledge: genome-wide
- From reductionist (1 gene) to the system (all genes)



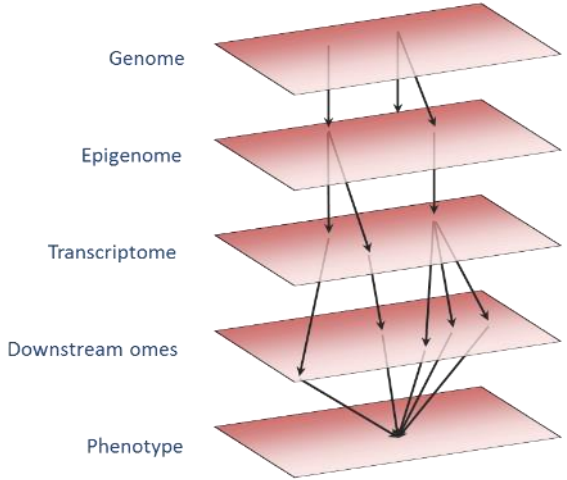
Acknowledges

- Complexity of biology
- Advantage of whole picture

Features of current biology

Data intensive

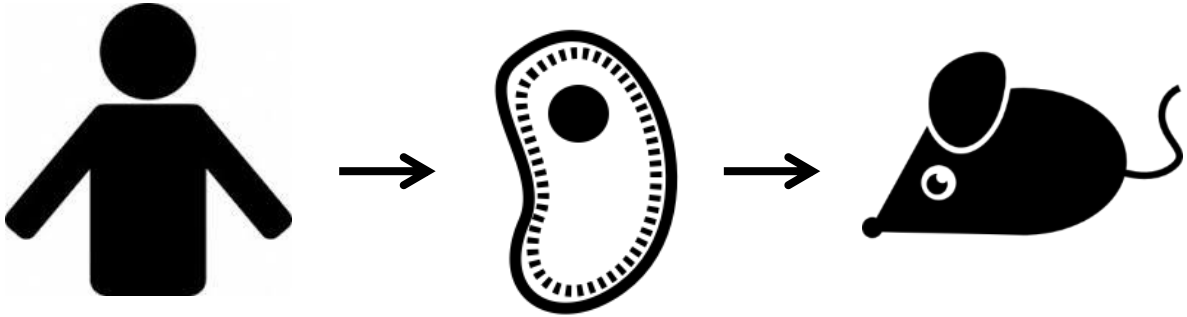
- Acknowledge lack of knowledge: genome-wide
- From reductionist (1 gene) to the system (all genes)
- Integrative: combine levels to trace processes



Features of current biology

Data intensive

- Acknowledge lack of knowledge: genome-wide
- From reductionist (1 gene) to the system (all genes)
- Integrative: combine levels to trace processes
- Using natural variation (instead of experimental) in large-scale population studies



Who

Tutors

- Ingrid Meulenbelt & Bas Heijmans (coordinators)
- 14 others

You

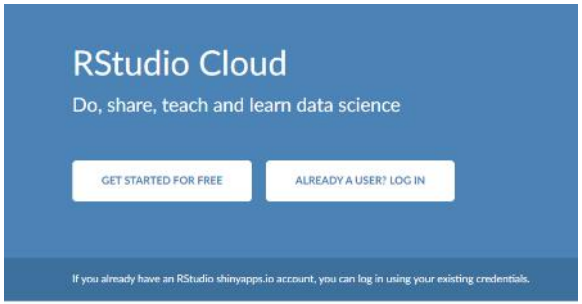
- 16 Master students Biomedical Sciences
- 2 PhD students (first 2.5 weeks)

The course

- Learning by doing: analyzing all kind of omics data → 12 days
- Applying knowledge: develop molecular data science project → 8 days

Tools

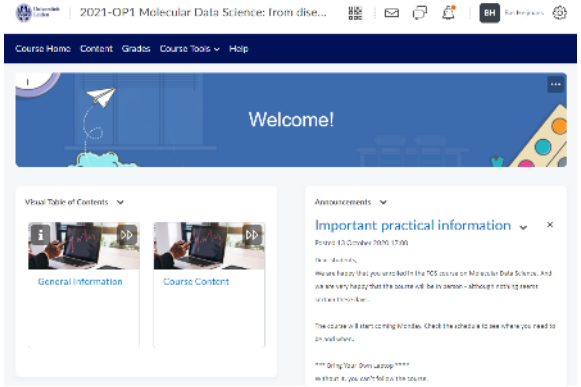
RStudio Cloud



Data science without the hardware hassles

Discord, Teams…

Brightspace



Learning objectives

Knowledge and understanding

- Knows how large-scale molecular data can inform on mechanisms and risk of common diseases.

- Has insight in modern data analysis methods used to discover molecular signatures of disease phenotypes in genetic, epigenetic, gene expression, and metabolomics data sets.

Applying knowledge and understanding

- Get hands-on experience in the analysis and interpretation of genetic, epigenetic, gene expression, and metabolomics data sets.
- Shows the ability to develop new researcher project in the field of ageing using molecular data science including background, hypothesis, pilot data, objectives, study design, work plan, and expected outcomes (e.g. causality).
- Can perform analyses to generate pilot data in order to critically appraise and, if necessary, reformulate a hypothesis.

Communication

- Shows communication skills to clearly and convincingly present and defend a research proposal.

- Is able to respond constructively to questions/feedback and connecting this feedback to his/her own position regarding his/her own research and in doing so showing an open, self-critical yet firm and self-confident attitude.

Learning skills

- Shows professional conduct: being critical yet constructive and eager to improve oneself and in doing so contributes to the learning process of the other students.

- Critically and constructively discusses research proposals of peers.

Pay close attention
Assessment

Assessment

1. Handing in assignments. Individually assessed (0%, P/F).
2. Contribute to interim evaluation of student participation and development during workgroups (0%).
3. Fill out project proposal form (for reflective assignment) (0%, P/F).
4. Presentation project proposal (background, hypothesis, pilot data, objectives, study design, workplan, expected outcomes) (45%).
5. Active and critical participation during discussion after project presentations of peers (15%).
6. Reflective assignment that shows mastering key aspects of development of research proposal in molecular data science and addressing points raised during peer review (40%).

Evaluation

Who will it be?

(2 are needed)

Every morning @10.30
Every afternoon @15.00

In the classroom

FOS course

Molecular Data Science

Part of the Master Track Data-driven Research